

## Regular article

# An improved semiempirical method for hydrated systems

W. Harb<sup>1</sup>, M. I. Bernal-Uruchurtu<sup>2</sup>, M. F. Ruiz-López<sup>1</sup>

<sup>1</sup> Equipe de Chimie et Biochimie théoriques, UMR CNRS-UHP No. 7565, Université Henri Poincaré-Nancy I, BP 239, 54506 Vandoeuvre-lès-Nancy, France

<sup>2</sup> Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, 62210 Cuernavaca, Mexico

Received: 28 June 2003 / Accepted: 27 August 2003 / Published online: 19 May 2004  
© Springer-Verlag 2004

**Abstract.** Hydrogen bonding is a key feature in many chemical and biochemical processes as well as in hydration phenomena. In previous works, it has been shown that description of hydrogen bonds by standard semiempirical methods is rather inaccurate and contains serious nonphysical artifacts. We have recently suggested a scheme to overcome this important limitation. Basically, we have proposed to replace the Gaussian correction function in the core-core interaction term of standard Austin model 1 (AM1) or parameterized model 3 (PM3) methods by a suitable function that is parameterized to correctly reproduce the long-range behavior of the interaction potential. A set of parameters for describing water-water interactions was derived using a highly refined *ab initio* potential energy surface for the water dimer. The present work extends that approach to the case of solute-solvent interactions of organic molecules in aqueous solution. Specifically, parameters for HH, HC, HN, HO, OC, ON and OO core-core terms are reported for use with the PM3 method. A series of tests for prototypical 1:1 complexes have been carried out and comparison with high-level *ab initio* results made. Errors in computed interaction energies are substantially smaller than those obtained in standard PM3 calculations, the root mean square decreasing from 3.41 to 1.74 kcal/mol. Artifacts in the potential energy surfaces (shoulders, spurious minima) present in PM3 are corrected with the new parameterization.

**Keywords:** Semiempirical methods – Parameterized model 3 – Hydrogen bond – Core-core interaction – Aqueous systems

## Introduction

In recent years, the affordability of quantum mechanical methods has made possible the detailed study of the structure and reactivity of many chemical and biochemical systems. However, for large systems, refined quantum computations are still too costly. The Car-Parrinello molecular dynamics methodology [1, 2] has opened a far-reaching approach but the state-of-the-art methodology and technology permits only the rigorous treatment of a narrow set of reactivity examples [3, 4, 5].

Indeed, the complexity of many interesting phenomena has, in general terms, forced the choice between two alternatives: the use of sophisticated *ab initio* methods that allow a refined treatment of a small model, or the use of approximate methods capable of providing a qualitative description of the whole system. Among them, the combined quantum mechanics and molecular mechanics (or simply QM/MM) techniques [6, 7, 8] have become very popular. A small-to-medium size part of the system is treated with a quantum method and the surroundings are incorporated by means of a classical treatment. However, in many cases, particularly in enzymatic catalysis studies, the quantum subsystem size is still too large to be treated at a high *ab initio* level and semiempirical methods have commonly been used within this QM/MM scheme [9].

Nowadays, a full quantum study of systems with several hundreds of atoms may be envisaged too, through the use of linear scaling algorithms. Again, semiempirical methods appear as a quite convenient choice and biological macromolecules have already been described at such a level [10, 11].

The essential features, strengths and weaknesses, of semiempirical theories have recently been reviewed by Clark [12]. Other overviews on semiempirical methods have been reported [13, 14, 15, 16, 17].

One of the most important shortcomings of primitive semiempirical methods was their inaccuracy in describing intermolecular interactions, in particular hydrogen bonds. Efforts were put forth to overcome this problem

and different models were developed. Among them, Austin model 1 (AM1) [18] and parameterized model 3 (PM3) [19] are the most general and have become the most used nowadays. Though some encouraging results were obtained with these methods with respect to the older ones [20, 21, 22], important deficiencies have been reported [23, 24, 25, 26, 27, 28]. Moreover, proton-transfer barriers occurring in hydrogen-bonded systems are generally overestimated by PM3 [29]. Studies by Trong-Anh et al. [30] concluded that AM1 and PM3 underestimate frontier interactions with respect to steric repulsion in some organic reactions. More recently, Cummins et al. [31] have made a decomposition analysis of the intermolecular interaction energy and have shown that the electrostatic component in semiempirical methods is mainly repulsive, whereas it accounts for a significant part of the stabilization energy in ab initio calculations. Spanget-Larsen [32] discussed the failure of neglect of diatomic differential overlap based methods to reproduce hydrogen bonds on the basis of an inadequate treatment of overlap effects and overestimation of nonbonded next-neighbor resonance integrals.

In addition to the problems already mentioned, some authors have demonstrated the presence of nonphysical artifacts in the computed potential energy surfaces by AM1 and PM3 [33, 34, 35, 36, 37] that are related to the presence of Gaussian correction functions (GCFs) in the core-core repulsion terms.

Some alternatives to standard AM1 and PM3 methods have been proposed; we mention only a few. An interesting computational strategy has been proposed by Gonzalez-Lafont et al. [38] in which the semiempirical method is reparameterized to fit ab initio data for a particular reacting system. The corresponding specific reaction parameters may be obtained using genetic algorithms [39]. Jug and Geudtner [40] used p polarization functions on hydrogen atoms to improve the hydrogen-bond description with SINDO/1. Svanberg et al. [41] obtained satisfactory results for neutral and protonated water clusters by using scaled PM3 GCFs. Other authors have proposed to correct the total PM3 energy by adding an empirical force field for the atoms involved in hydrogen bonds [42], in a way similar to the MNDO/H [43] and MNDO/M [44] procedures. Martin [45] has suggested a self-consistent correction to account for dispersion contributions by adding London terms to the Fock matrix. Repasky et al. [46] have looked at the substantial error in semiempirically calculated heats of formation of organic molecules and proposed a pairwise distance-directed Gaussian (PDDG) approach. A PDDG function is added to the core repulsion function of either PM3 or modified neglect of differential overlap (MNDO) and a whole reparameterization of the corresponding PDDG-PM3 and PDDG-MNDO methods is carried out. Mean absolute errors in calculated heats of formation of single molecules are significantly reduced with respect to the parent method. For intermolecular interactions, PDDG-MNDO displays similar behavior to MNDO, while PDDG-PM3 improves the results of PM3. However, some stationary structures of the water dimer could not be located. The authors also noted that

spurious minima introduced by PM3 core-repulsion functions are retained in PDDG-PM3.

We have recently proposed a scheme that considerably improves the description of hydrogen-bonded systems [47, 48]. Basically, the conventional PM3 GCFs are replaced by suitable functions exhibiting the correct physical behavior in the whole range of intermolecular separation distances. In the simplest approach, parameterized interaction functions (PIFs) are introduced for atoms belonging to different molecules [47]. The parameters of these PIFs are optimized to reproduce high-level ab initio results for the intermolecular potential energy surface (IPES) of reference systems. Promising results were obtained for the water dimer and water clusters with such a PM3-PIF approach. The PIFs may then be used to derive a single core-core function utilizable in the whole range of interatomic distances (i.e., without differentiating intra molecular and intermolecular terms). A possible method to achieve such a goal has been proposed and it has been shown to describe pretty well proton transfer in the protonated water dimer [47] and water clusters (M.I. Bernal-Uruchurtu, M.F. Ruiz-López MF, unpublished results) as well as the structure of liquid water (G. Monard, M.I. Bernal-Uruchurtu, A. Van der Vart, K.M. Merz Jr., M.F. Ruiz-López, unpublished results). Improvements with respect to standard PM3 results were remarkable.

Encouraged by these preliminary studies on aqueous systems, our aim has been to extend the underlying idea to optimize the core-core functions of other atom types. We report here PIF parameters for HH, HC, HN, HO, OC, ON and OO core-core interactions. These functions may allow us, in particular, to achieve a reliable description of systems containing H, N, C and O atoms interacting with water molecules at the PM3 level. Therefore, the present results are potentially useful to many PM3 users, particularly for those working on the modeling of enzymatic reactions or processes in liquid water.

Before describing in detail our work, we illustrate PM3 artifacts in the next section. Afterwards, the parameterization procedure is described and tests for the energetics and structure of some hydrogen-bonded systems will be discussed.

### Artifacts in PM3 intermolecular interaction energies

The MNDO method [49] was not capable of predicting hydrogen-bonded structures owing to its propensity to overestimate repulsion. AM1 [18] and PM3 [19] are both based on the MNDO method but use a different formulation of the core-core repulsion between atoms A and B which is obtained as

$$E_{AB}^{\text{core}} = Z_A Z_B \langle S_A S_A | S_B S_B \rangle (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) + g(A, B). \quad (1)$$

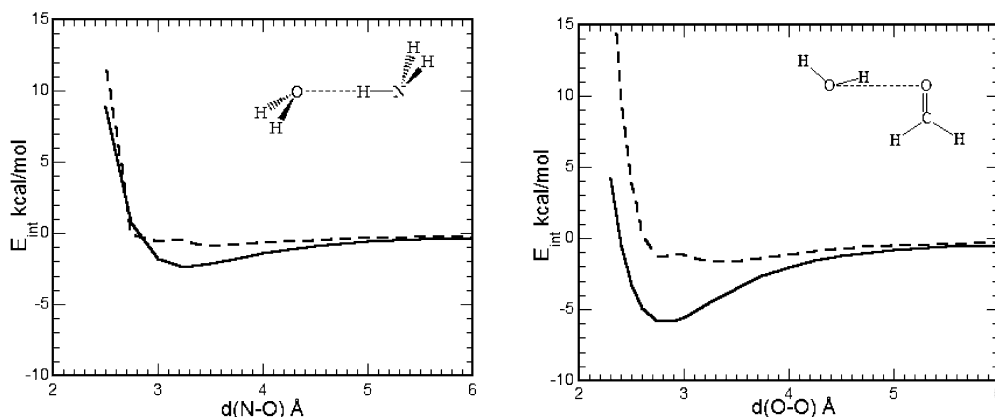
Here,  $Z_A$  and  $Z_B$  are the core charges of atoms A and B, respectively,  $S_A$  and  $S_B$  are s atomic orbitals,  $R_{AB}$  is the internuclear distance,  $\alpha_A$  and  $\alpha_B$  are atomic parameters

and  $g(A,B)$  is a correction function absent in the MNDO method. Note that for the case when  $A$  is oxygen or nitrogen and  $B$  is hydrogen, the leading exponential term is multiplied by  $R_{AH}$ . For both AM1 and PM3 methods,  $g(A,B)$  involves a sum of atomic GCF terms as

$$g(A,B) = \frac{Z_A Z_B}{R_{AB}} \left( \sum_i K_{A,i} e^{-L_{A,i}(R_{AB}-M_{A,i})^2} + \sum_j K_{B,j} e^{-L_{B,j}(R_{AB}-M_{B,j})^2} \right). \quad (2)$$

In this expression,  $K$ ,  $L$  and  $M$  are adjustable parameters that modulate the Gaussian function. In the  $g(A,B)$  parameterization procedure, the interaction energy and/or the best-known available structure for a few hydrogen-bonded complexes were taken into account. Nevertheless, the experimental information was rather scarce and the approach soon proved to have severe limitations. Many works have described these problems and a few of them were referenced in the Introduction. In this section, we just illustrate some distinctive artifacts of the PM3 method in the treatment of intermolecular interactions.

Ab initio and PM3 potential energy curves for two 1:1 complexes are shown in Fig. 1. These curves correspond to the interaction energies defined as  $\Delta E = E(\text{complex}) - \Sigma E(\text{monomers})$ . The monomers are assumed to have the experimental geometry and their structure is not allowed to change in the complex. Interaction energies are computed as a function of intermolecular separation either at the MP2/aug-cc-pVQZ (without  $g$  functions) level (ab initio curve) or at the PM3 level. Configurations of the complex were generated in the following way. We first carried out a geometry optimization of the complex at the MP2/6-311++(d,p) level with the internal geometry of the monomers constrained to their experimental structure. The  $X$ - $Y$  interatomic distance was then scanned (all other degrees of freedom being fixed) to build the energy curve.



**Fig. 1.** Interaction energy curves for some 1:1 complexes. Comparison between ab initio MP2/aug-cc-pVQZ (without  $g$  functions) (solid line) and parameterized model 3 (PM3) (dashed line) results. The monomers are assumed to have the experimental geometry and their structure is not allowed to change in the complex. Configurations of the complex were generated in the following way. We first

Though these curves are somewhat arbitrary, they clearly point out two common problems with PM3. First, stabilization by hydrogen bonding of stationary structures is often underestimated compared with ab initio data, and, second, shoulders and spurious energy minima appear in the curves.

Another typical PM3 default is shown in Fig. 2 where the fully optimized structures of some problematical 1:1 complexes are represented. All complexes exhibit an interatomic H-H distance about 1.7–1.8 Å and an interaction energy around 2 kcal/mol. Obviously, this is a nonphysical result since such structures are clearly repulsive owing to the close H-H distance. As discussed later and first reported by Buss et al. [33], this artifact is directly related to the hydrogen-hydrogen GCF in PM3 that introduces an artificial minimum at that distance range. This fact considerably limits the applicability of PM3 to describe hydrogen bonds and also steric interactions. Indeed, PM3 (and AM1 as well) has been shown to fail in asymmetric synthesis studies, because the most hindered structures are predicted to be the stablest ones, owing to H-H stabilizing interactions [37].

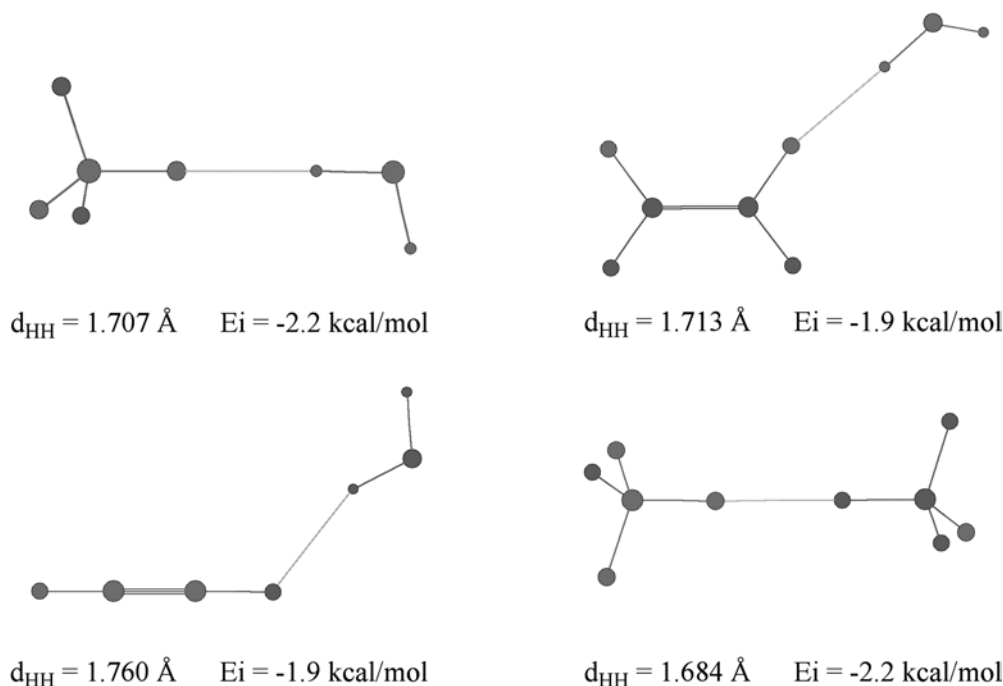
### The PIF approach

The problems just described encourage the search of adequate core-core functions parameterized to describe the correct physics of nonbonded interactions. The PIF that we have proposed before [48] is built up as a sum of atom-atom contributions in close relation to those used in effective potentials:

$$\text{PIF} = \sum_{A,B}^{\text{inter}} g(A,B) = \sum_{A,B}^{\text{inter}} \left( \alpha_{AB} e^{-\beta_{AB} R_{AB}} + \frac{\chi_{AB}}{R_{AB}^6} + \frac{\delta_{AB}}{R_{AB}^8} + \frac{\epsilon_{AB}}{R_{AB}^{10}} \right), \quad (3)$$

where  $\alpha_{AB}$ ,  $\beta_{AB}$ ,  $\chi_{AB}$ ,  $\delta_{AB}$  and  $\epsilon_{AB}$  are adjustable parameters depending on  $(A,B)$  atom types. Each

carried out a geometry optimization of the complex at the MP2/6-311++(d,p) level with the internal geometry of the monomers constrained to their experimental structure. The  $X$ - $Y$  interatomic distance was then scanned (all other degrees of freedom being fixed) to build the energy curve



**Fig. 2.** PM3 fully optimized structures for  $\text{CH}_4\text{-H}_2\text{O}$ ,  $\text{C}_2\text{H}_4\text{-H}_2\text{O}$ ,  $\text{C}_2\text{H}_2\text{-H}_2\text{O}$  and  $\text{CH}_4\text{-CH}_4$  complexes. Notice the nonphysical short H-H distances

$g(A,B)$  replaces the corresponding GCF for the  $(A,B)$  intermolecular pair in the AM1 or PM3 nuclear repulsion function. The exponential term may be related to short-range repulsion, whereas the remaining terms may be essentially related to electrostatic plus induction plus dispersion corrections to the semiempirical intermolecular energy. However, in the present context, the PIF is merely an auxiliary function necessary to fit a set of reference data, so a formal interpretation of the optimized coefficients should not be attempted. The parameters are obtained by fitting high-level ab initio results for the IPES of reference systems, as explained in the following.

### Parameterization procedure

In the present work, our aim has been to derive a set of parameters that allows PM3 studies on hydration or aqueous solvation phenomena. In other words, we look for PIF functions that improve the standard PM3 description of a molecule interacting with one or several surrounding water molecules.

As previously, PIF functions are obtained by fitting some high-level reference ab initio IPES [48]. Here, we present a set of parameters for molecules containing H, C, N and/or O atoms. The reference IPESs correspond to the following 1:1 complexes: water dimer, methane-water, ammonia-water, formaldehyde-water, methanol-water, formic acid-water and hydrocyanic acid-water. The configurations present in the energy surfaces explore different orientations of the molecules and sample the attractive and repulsive regions of the IPES. The number of configurations considered for each complex

was 658 for water-water; 293 for water-ammonia; 377 for methane-water, 69 for formaldehyde-water, 19 for methanol-water, 21 for formic acid-water and 47 for hydrocyanic acid-water (total number 1,484). It should be noted that the geometry of the monomer always corresponds to the experimental one. Ab initio interaction energies for these systems were obtained at the MP2 level using the aug-cc-pVQZ (without g functions) basis set. Owing to the size of the basis set, the basis set superposition error can be safely neglected [50]. The water-water IPES was obtained before [51] and the IPES for the other systems were computed in the present work. The ab initio calculations reported in this paper were performed using Gaussian 98 [52], whereas the semiempirical computations were done using the GEOMOS package [53].

The search for suitable parameters for the seven different  $A$ - $B$  atom pair interactions (HO, CO, NO, OO, HH, CH and NH) was done using an ad hoc optimization scheme that simultaneously fitted all the IPESs for the previously described systems. We present the best set of parameters found in Table 1.

The PIF functions appear in Fig. 3, where they are compared with the original PM3 GCFs functions. In contrast with the latter, PIF functions have a nonnegligible contribution in the region corresponding to intermolecular interactions. Additionally, it can be seen that all of them exhibit smooth behavior. Curves for  $XH$  pairs display a minimum and give a negative (stabilizing) contribution to the interaction energy at typical hydrogen-bond distances. The depth of the well decreases in the order  $\text{NH} > \text{OH} > \text{CH}$ . Conversely, HH and  $XY$  curves displays repulsive behavior for the whole distance range.

**Table 1.** Optimized parameters for parameterized interaction functions (PIFs). All parameters are in hartrees and bohrs

$A-B$	$\alpha_{AB}$	$\beta_{AB}$	$\chi_{AB}$	$\delta_{AB}$	$\epsilon_{AB}$
O-O	75.15466	1.512063	338.656	-32,185.68	274,634.10
O-N	1.94702	1.016194	1.941	35.45	887.74
O-C	0.87884	0.886032	1.879	33.95	843.67
O-H	29.32517	2.092648	-55.779	44.53	313.44
N-H	33.11381	1.782993	-172.487	158.43	2,423.97
C-H	0.42691	0.765030	-377.328	4,668.31	-7,005.27
H-H	2.47949	2.896120	47.262	-505.73	1,549.81

The final set of parameters was thoroughly checked to ensure that no artifacts were present. In the following sections, we present an analysis of the results obtained with the standard PM3 and the improved PM3-PIF for a series of hydrogen-bonded systems.

### Interaction energies for the training set

We show the correlation between the PM3 (standard, corrected) and ab initio interaction energies for the whole set of 1,484 reference configurations in Fig. 4. As can be seen, the corrected PM3-PIF is able to predict IPESs in much better agreement with ab initio data. The correlation coefficient significantly increases from 0.760 (PM3) to 0.900 (PM3-PIF). The slope of the regression line is remarkably ameliorated too (from 0.851 to 1.005).

In order to illustrate the improvement of PIF parameters for specific hydrogen bonds, some interaction energy curves are shown in Fig. 5. The curves were obtained in the same way as those in Fig. 1. The PM3-PIF profiles display a good to very good agreement with ab initio data and do not exhibit any of the artifacts (spurious minima, shoulders) of the original PM3 method. Look, for instance, at the curves for methane, acetonitrile and formaldehyde.

The differences found with respect to ab initio calculations, concerning either the position of the minima or the depth of the wells, are due in part to the use of constrained geometries. As a matter of fact, geometry relaxation tends to improve the agreement, as shown later.

### Interaction energies and optimized structures of test complexes

As a further test, the geometry of a set of 1:1 complexes with water was fully optimized. The test set includes all the molecules used in the fitting procedure (water, methane, ammonia, formaldehyde, methanol, formic acid and hydrocyanic acid) as well as other small molecules (acetylene, dimethyl ether, ethylene, formamide, formalimine, acetonitrile, acetone).

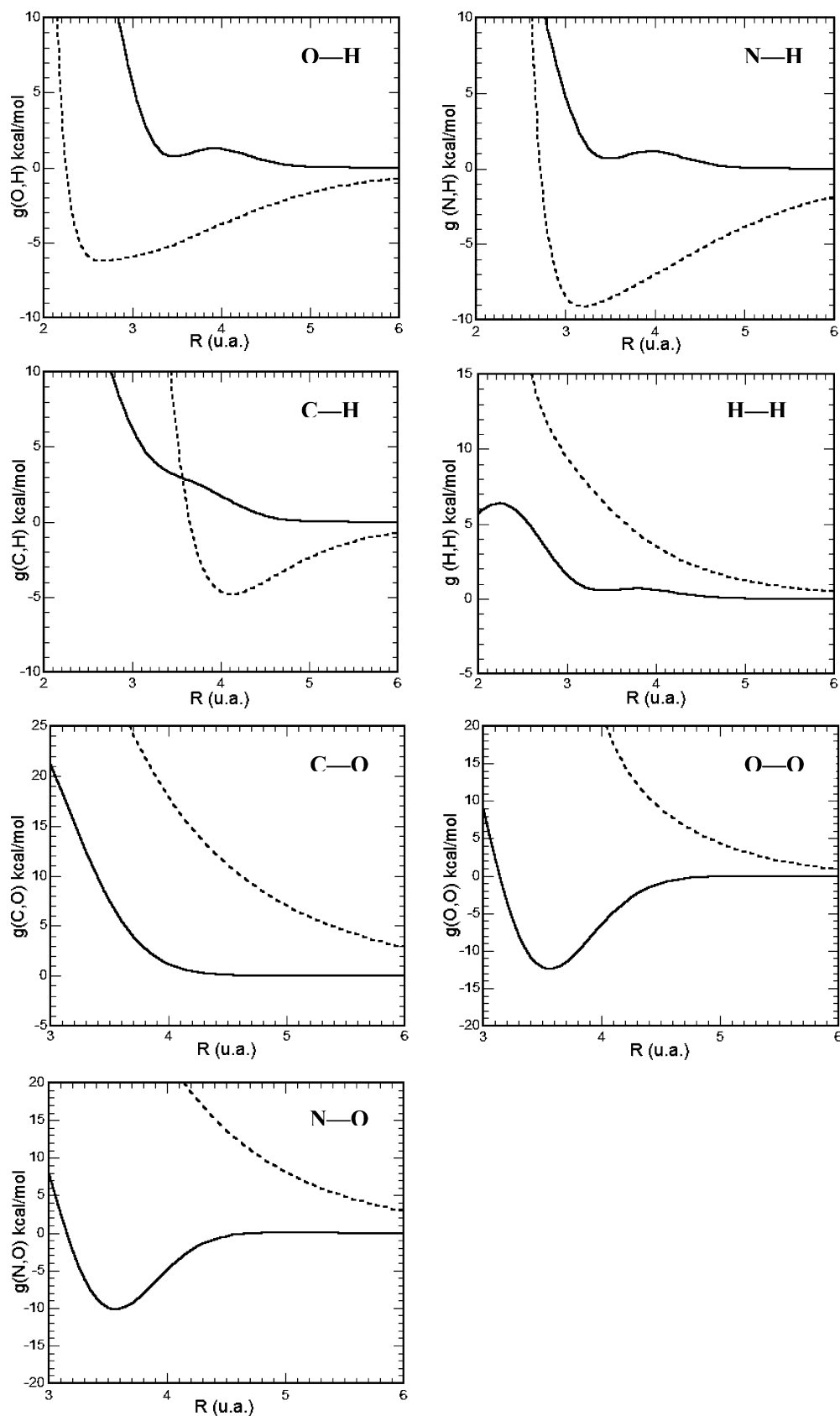
The ab initio results were obtained as follows. The optimized structure for the complexes was computed at the MP2/6-311++G(d,p) ab initio level. Then, the interaction energy was calculated at the MP2/aug-cc-pVQZ level (without g functions). Afterwards, PM3 and PM3-PIF geometry optimizations were carried out. It should be noted that for each complex several configurations may exist.

We did not make a systematic study of all these structures; instead, we considered the prototypical configurations for proton donors and proton acceptors, and for two cases (formamide and formic acid), the structures involving cyclic coordination.

The structures obtained are represented in Fig. 6. The ab initio and semiempirical interaction energies are compared in Table 2, while some structural parameters are compared in Table 3.

First of all, a few comments have to be made for the standard PM3 results. It must be noted that the PM3 energy minima for methane, ethylene and acetylene in Fig. 6 do not correspond to global minima. The stablest structures for these complexes correspond to those having close HH interaction, as shown in Fig. 2. This kind of configuration may be found for other complexes such as  $\text{NH}_3(\text{donor})-\text{H}_2\text{O}$  ( $E_{\text{int}} = -1.05$  kcal/mol) or hydrocyanic acid (acceptor)- $\text{H}_2\text{O}$  ( $E_{\text{int}} = -1.20$  kcal/mol). Besides, it is important to mention that for a given complex and for a given interaction arrangement type, PM3 may lead to several stable structures with different intermolecular distances owing to the presence of multiple minima in the IPES. However, for sake of comparison with ab initio results, we have only considered structures of equivalent type. One must also note that PM3 gives a repulsive interaction energy for the  $\text{CH}_4-\text{H}_2\text{O}$  complex (Table 2). This can be understood by looking at the corresponding potential energy curve in Fig. 5, where a minimum at about  $R = 3$  au, with positive energy, is presented (note that this curve was obtained with constrained monomer geometries). Finally, Table 3 shows all the optimized hydrogen-bond distances are too short in PM3 calculations, with an  $X \cdots H$  length of about 1.8 Å for all cases. The PM3 GCFs for NH and OH interactions (Fig. 5) have an energy minimum around the 1.7–1.8-Å region (i.e., about 3.5 au) and were shown to be of considerable influence for the geometry of the water dimer [36]. Our calculations reveal that this fact is rather systematic.

Several interesting trends are also shown in Table 2. Standard PM3 calculations systematically underestimate interaction energies. The largest PM3 error appears for the cyclic formamide-water complex, which amounts for -6.90 kcal/mol, whereas the largest PM3-PIF error happens for hydrocyanic acid (donor) and is -4.32 kcal/mol. In general, the PM3-PIF parameterization leads to better interaction energies than standard PM3, the absolute error being drastically smaller in several cases. The PM3-PIF root-mean-square (RMS) error is about half the PM3 one. Note that geometry optimization at



**Fig. 3.** Core-core correction functions,  $g(A,B)$ , as a function of internuclear separation,  $R$ , for a series of  $A-B$  atoms. Standard PM3 Gaussian correction functions (Eq. 2) (solid lines); parameterized interaction functions proposed in this work (Eq. 3) (dotted lines)

the semiempirical level (PM3 or PM3-PIF) decreases the RMS error, as advanced earlier.

It may be interesting to note that, for the case of proton donors, PM3-PIF interaction energies are always smaller than the *ab initio* ones (in absolute value), whereas for some proton acceptors, PM3-PIF tends to overestimate interaction energies.

Comparison of interaction energies obtained with the present PM3-PIF parameterization might also be done with the corresponding values reported in the recent work of Repasky et al. [46] using the PDDG-PM3 approach. The case of the water dimer was discussed in some detail by these authors. For the stablest configuration, they reported an interaction energy of  $-3.72$  kcal/mol, which is slightly closer to the *ab initio* result ( $-5.04$  kcal/mol at the level used in our work) than the standard PM3 calculation ( $-3.50$  kcal/mol). The improvement obtained with our parameterization ( $-4.51$  kcal/mol) is, therefore, substantially larger. This is not surprising considering that our approach was specifically developed to describe hydrated systems, whereas the PDDG/PM3 method was parameterized with the main aim of improving molecular heats of formation.

Let us now compare the structural quantities in Table 3. PM3-PIF leads, on average, to better geometries than standard PM3, the RMS errors for  $X\cdots H$  and  $X\cdots Y$  distances and for  $X\cdots Y\cdots H$  angles being smaller for the reparameterized method.

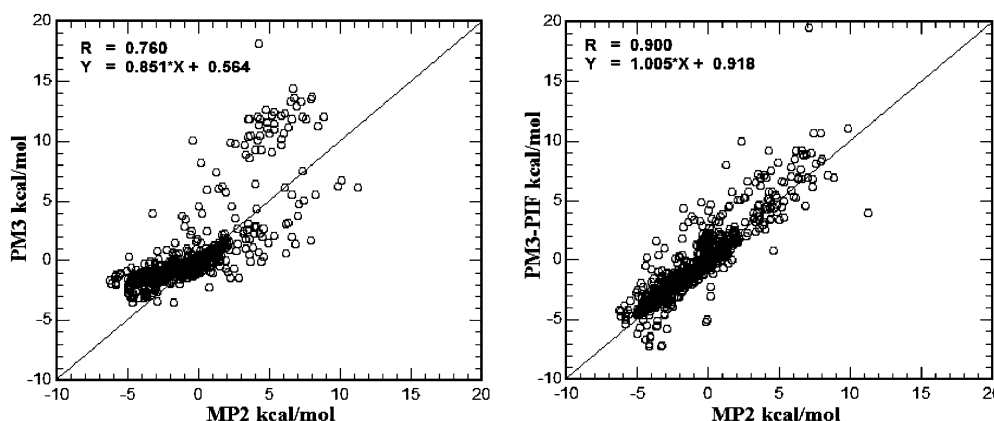
The main error of the PM3-PIF approach happens for the cyclic complexes and in particular for the formamide-water complex. Indeed, if these structures are excluded, the RMS errors for hydrogen bonds  $X\cdots H$  become 0.302 for PM3 and 0.165 for PM3-PIF. The RMS errors for  $X\cdots Y$  distances become 0.291 for PM3 and 0.150 for PM3-PIF. Cyclic complexes are particularly difficult to describe because a subtle balance of the interactions involved is needed. In the case of formic acid, for example, the PM3-PIF calculation leads to a structure displaying quite asymmetric hydrogen bonds. The length of the  $CO\cdots H^W$  bond is

$1.733$  Å, whereas the length of the  $OH\cdots O^W$  bond is  $2.385$  Å. This may be understood by considering the energy data in Table 2 and the structural parameters in Table 3 for methanol-water and formaldehyde-water, which may be considered as prototypical systems for  $OH\cdots O^W$  and  $CO\cdots H^W$  bonds. The PM3-PIF hydrogen-bond strength is underestimated for methanol (donor), whereas it is overestimated for formaldehyde. In contrast, standard PM3 underestimates the energy of both types of hydrogen bonds and predicts lengths that are too short in both cases. Hence, the later method leads to a cyclic formic acid-water complex with two hydrogen bonds of similar length, close to  $1.8$  Å.

## Conclusions

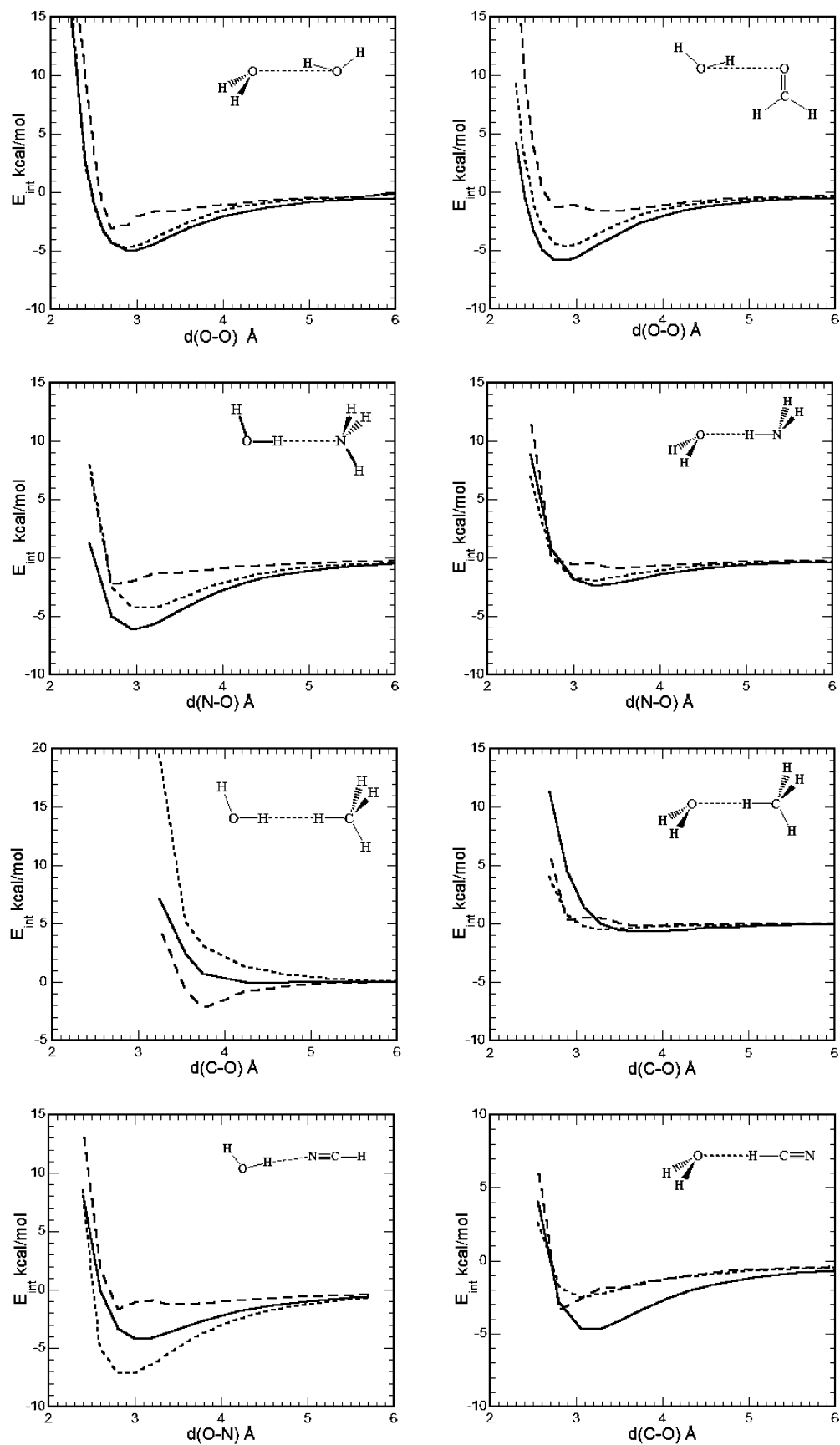
The results presented in this paper suggest that the PM3-PIF parameterization is suitable for treating molecular systems in interaction with water molecules. Compared with the standard method, the PM3-PIF approach leads to a better agreement with high-level *ab initio* results for both interaction energies and geometrical parameters. The RMS error for interaction energies is substantially decreased (by a factor of about 2) and predicted structures are in better agreement with *ab initio* ones. Moreover, the artifacts present in the standard PM3 method are completely removed. In particular, the defective structures due to H-H stabilizing interactions such as those obtained for the methane-water complex are no longer expected.

The parameters reported in this work are limited to molecules containing H, C, N and O atoms but other interaction types may be parameterized following a similar procedure and work is under progress. Finally, it must be underlined that the present approach is limited to a reparameterization of GCFs in the core-core repulsion function. One could envisage a whole reparameterization of the PM3 (or AM1) method in which the GCFs are replaced by "physically appropriate" func-



**Fig. 4.** Comparison between semiempirical and *ab initio* interaction energies for the complexes used in the PM3-parameterized interaction function (PIF) parameterization. The number of configurations considered for each complex was 658 for water-water; 293

for water-ammonia, 377 for methane-water, 69 for formaldehyde-water, 19 for methanol-water, 21 for formic acid-water and 47 for hydrocyanic acid-water (total number 1,484)



**Fig. 5.** Comparison between ab initio (solid line), PM3 (dashed line), and PM3-PIF (dotted line) interaction energy curves of some hydrogen-bonded systems. The geometry of the monomers is the experimental one



### Proton donors


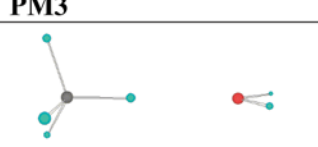


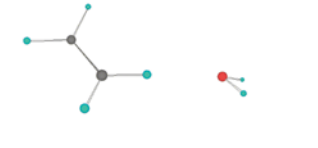

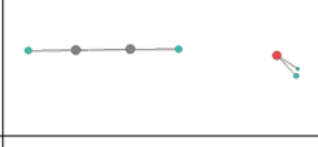








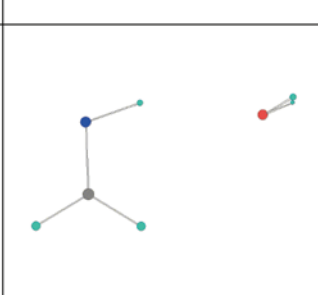
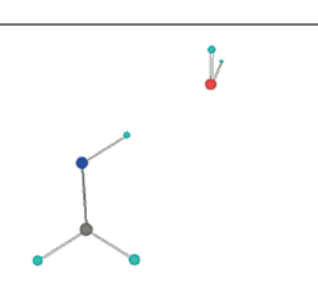
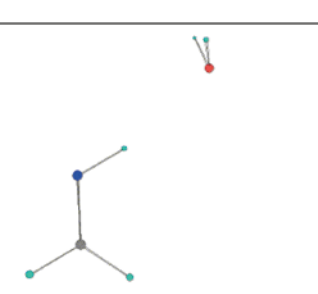
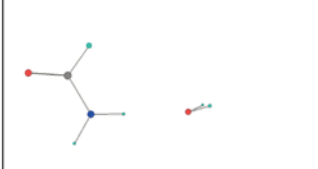
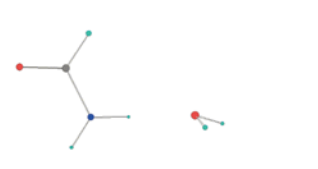
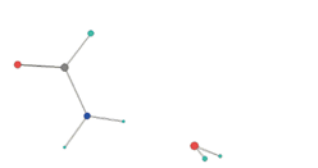



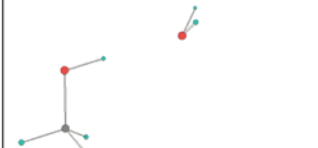
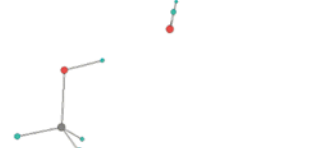
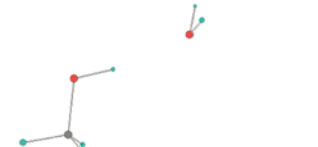
	MP2	PM3	PM3-PIF
CH <sub>4</sub>			
CH <sub>2</sub> CH <sub>2</sub>			
HCCH			
HCN			
NH <sub>3</sub>			
CH <sub>2</sub> NH			
NH <sub>2</sub> COH			
H <sub>2</sub> O			
CH <sub>3</sub> OH			

Fig. 6a-c. Comparison between ab initio, PM3 and PM3-PIF fully optimized structures for selected test cases

### Proton acceptors

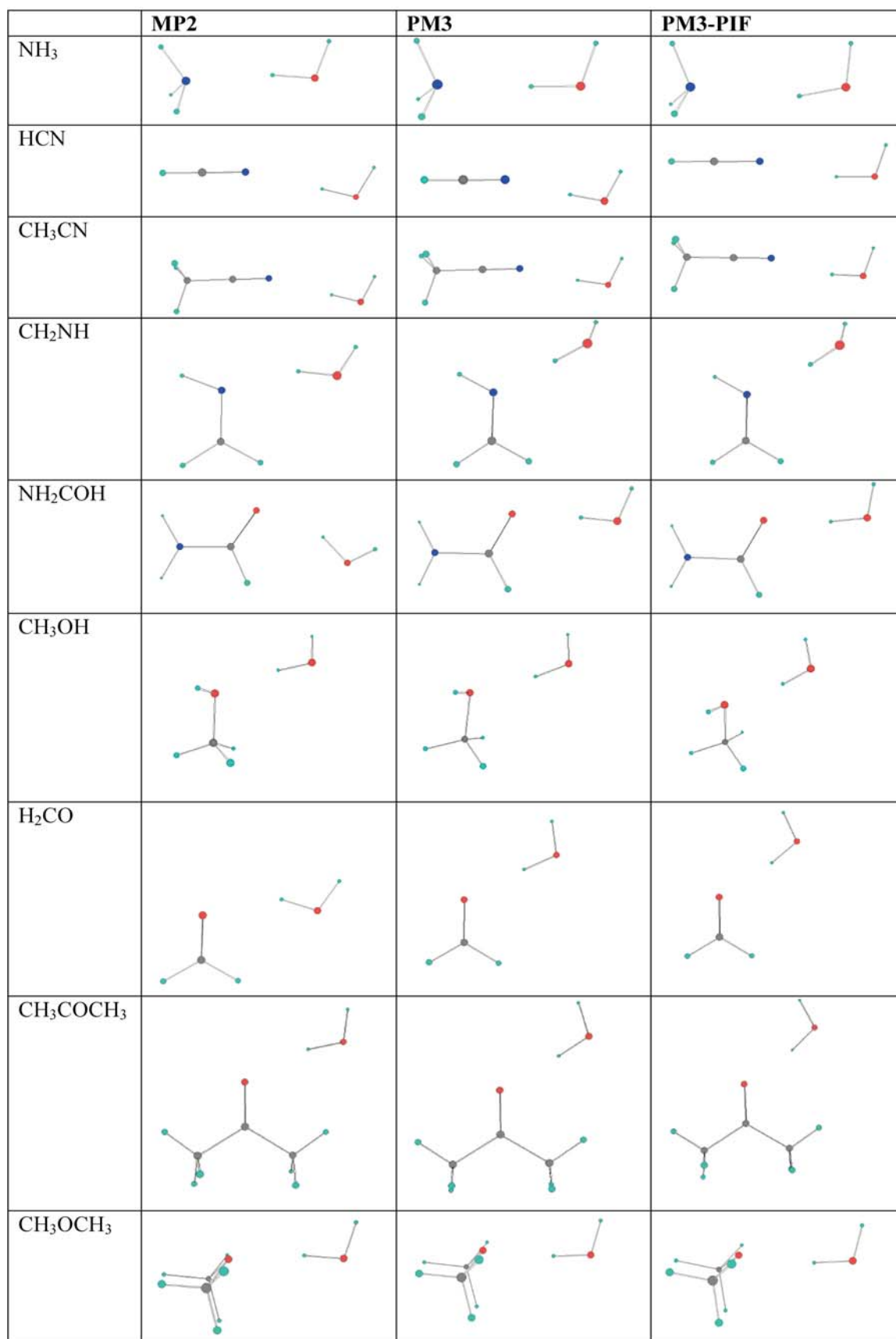


Fig. 6a-c. (Continued)

## Cyclic dimers

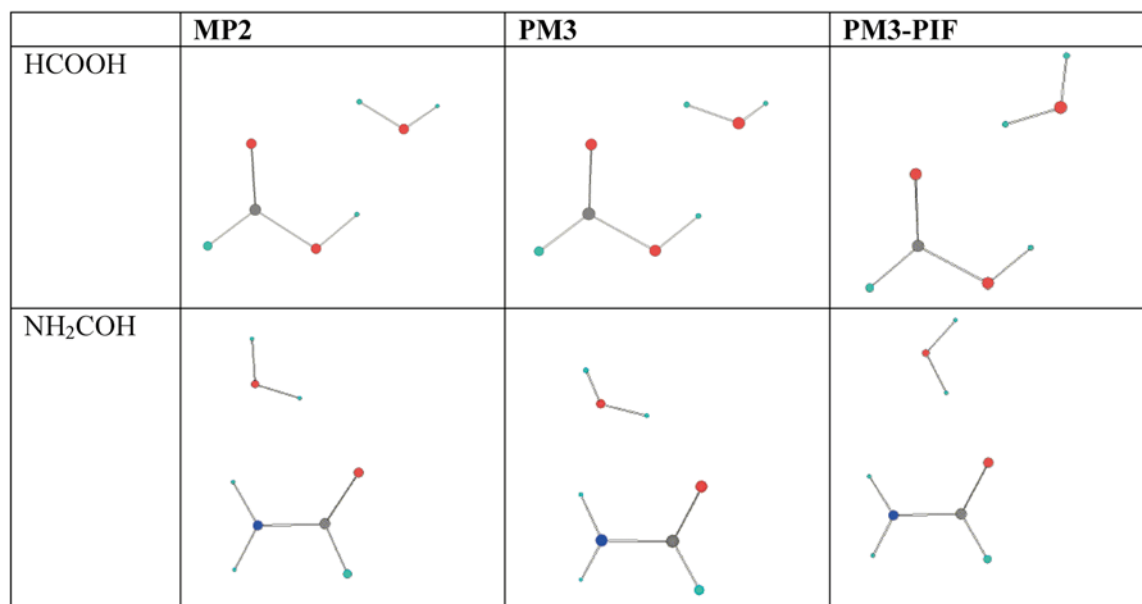


Fig. 6a-c. (Continued)

**Table 2.** Comparison of interaction energies for 1:1 complexes with water (kcal/mol)

	MP2 <sup>a</sup>	PM3		PM3-PIF			
		single-point calculation <sup>b</sup>	Optimized geometry <sup>c</sup>	Energy difference <sup>d</sup>	single-point calculation <sup>b</sup>	Optimized geometry <sup>c</sup>	Energy difference <sup>d</sup>
H donors							
CH <sub>4</sub>	-0.64	-0.11	+0.05	-0.69	-0.33	-0.61	-0.03
C <sub>2</sub> H <sub>2</sub>	-1.14	-0.18	-0.48	-0.66	-0.22	-0.49	-0.65
HCCH	-2.94	-0.60	-1.99	-0.95	-1.15	-1.32	-1.62
HCN	-7.73	-5.41	-4.20	-2.32	-6.24	-3.41	-4.32
NH <sub>3</sub>	-2.37	-0.27	-0.88	-1.49	-1.90	-1.94	-0.43
CH <sub>2</sub> NH	-3.30	-0.20	-1.50	-1.80	-0.44	-1.73	-1.57
NH <sub>2</sub> COH	-5.54	-0.61	-2.21	-3.33	-1.40	-2.30	-3.24
H <sub>2</sub> O	-5.04	-2.68	-3.50	-1.54	-4.6	-4.51	-0.53
CH <sub>3</sub> OH	-5.11	-2.54	-3.21	-1.90	-3.35	-3.37	-1.74
H acceptors							
NH <sub>3</sub>	-6.56	-2.17	-3.04	-3.52	-4.38	-4.63	-1.93
HCN	-6.67	-3.85	-1.73	-4.94	-9.67	-7.44	0.77
CH <sub>3</sub> CN	-5.09	-1.21	-2.01	-3.08	-7.37	-7.84	2.75
CH <sub>2</sub> NH	-6.80	-1.65	-2.35	-4.45	-5.92	-6.67	-0.13
NH <sub>2</sub> COH	-7.02	-2.66	-3.75	-3.27	-6.91	-8.67	1.65
H <sub>2</sub> O	-5.04	-2.68	-3.50	-1.54	-4.60	-4.51	-0.53
CH <sub>3</sub> OH	-5.86	-2.35	-2.92	-2.94	-4.43	-4.76	-1.10
H <sub>2</sub> CO	-5.52	-1.32	-2.74	-2.78	-4.78	-6.77	1.25
(CH <sub>3</sub> ) <sub>2</sub> CO	-6.84	-2.75	-3.63	-3.21	-6.30	-7.54	0.70
(CH <sub>3</sub> ) <sub>2</sub> O	-6.22	-2.40	-2.50	-3.72	-4.96	-5.08	-1.14
Cyclic structures							
HCOOH	-10.27	-4.38	-5.77	-4.5	-6.03	-8.51	-1.76
NH <sub>2</sub> COH	-9.9	-1.55	-3.93	-6.90	-6.34	-9.46	-0.44
Root-mean-square error		3.98	3.41		2.12	1.74	

<sup>a</sup> MP2/ aug-cc-pVQZ (without g functions) interaction energies using optimized MP2/6-311++G\*\* geometries

<sup>b</sup> Semiempirical single-point energy calculations using MP2/6-311++G\*\* optimized geometries

<sup>c</sup> Interaction energies with optimized structures at the semiempirical level (PM3 or PM3-PIF)

<sup>d</sup> Energy difference between MP2 and semiempirical interaction energies

**Table 3.** Comparison of geometrical parameters for selected hydrogen-bonded complexes. MP2 calculations were carried out with the 6-311++G\*\* basis set. All distance values are in angstroms, angles are in degrees

	$X\cdots H$ distance			$X\cdots Y$ distance			$X-Y-H$ angle		
	MP2	PM3	PIF	MP2	PM3	PIF	MP2	PM3	PIF
H donors									
CH <sub>4</sub>	2.551	1.852	2.200	3.64	2.949	3.294	1.3	0.2	0.3
C <sub>2</sub> H <sub>2</sub>	2.444	1.844	2.310	3.515	2.942	3.401	7.5	1.8	3.0
HCCCH	2.194	1.824	2.122	3.262	2.903	3.198	2.6	2.4	0.2
HCN	2.064	1.798	1.920	3.139	2.885	3.009	0.0	0.0	0.0
NH <sub>3</sub>	2.207	1.866	2.168	3.222	2.867	3.161	2.3	5.7	7.4
CH <sub>2</sub> NH	2.188	1.850	2.265	3.144	2.850	3.240	17.3	0.6	9.5
NH <sub>2</sub> COH	2.001	1.838	2.218	3.012	2.839	3.218	0.9	4.2	2.3
H <sub>2</sub> O	1.95	1.809	1.928	2.914	2.769	2.882	2.2	1.0	5.4
CH <sub>3</sub> OH	1.943	1.812	2.026	2.908	2.761	2.970	0.7	6.8	7.6
H acceptors									
NH <sub>3</sub>	1.976	1.825	2.087	2.941	2.792	3.035	5.7	1.2	8.0
HCN	2.16	1.846	1.909	3.124	2.809	2.868	0.7	1.5	7.0
CH <sub>3</sub> CN	2.106	1.842	1.896	3.071	2.806	2.855	0.3	2.0	7.3
CH <sub>2</sub> NH	1.971	1.834	1.919	2.901	2.800	2.878	13.5	1.1	7.0
NH <sub>2</sub> COH	1.93	1.807	1.732	2.838	2.768	2.705	16.6	1.9	3.3
H <sub>2</sub> O	1.95	1.809	1.928	2.914	2.769	2.882	2.2	1.0	5.4
CH <sub>3</sub> OH	1.896	1.819	1.873	2.857	2.773	2.830	5.2	5.0	6.0
H <sub>2</sub> CO	2.017	1.817	1.783	2.889	2.777	2.751	21.0	1.9	3.3
(CH <sub>3</sub> ) <sub>2</sub> CO	1.94	1.812	1.740	2.882	2.773	2.710	11.0	1.9	5.0
(CH <sub>3</sub> ) <sub>2</sub> O	1.864	1.827	1.821	2.828	2.784	2.786	5.0	4.3	4.7
Cyclic structures									
HCOOH									
OH $\cdots$ O <sup>w</sup>	1.792	1.78	2.385	2.733	2.679	3.180			
CO $\cdots$ H <sup>w</sup>	2.142	1.823	1.733	2.833	2.618	2.698			
NH <sub>2</sub> COH									
NH $\cdots$ O <sup>w</sup>	2.063	1.856	2.739	2.898	2.738	3.428			
CO $\cdots$ H <sup>w</sup>	1.958	1.828	1.721	2.818	2.663	2.697			
Root-mean-square error		0.286	0.263		0.272	0.204		8.2	6.8

tions, provided enough information on the potential energy surfaces for interacting molecules is included in the data set to be fitted. Our PM3-PIF approach gives a few indications on how such a task could be accomplished.

*Acknowledgements.* The authors acknowledge financial support from the Ecos-Nord Program (project no. M98-P01) and the CNRS computer center CINES (project lct2550) for computational facilities.

## References

- Car R, Parrinello M (1985) *Phys Rev Lett* 55:2471–2474
- Lippert G, Hutter J, Parrinello M (1999) *Theor Chem Acc* 103:124–140
- Rousseau R, Boero M, Bernasconi M, Parrinello M, Terakura K (2000) *Phys Rev Lett* 85:1254–1257
- Mundy CJ, Hutter J, Parrinello M (2000) *J Am Chem Soc* 122:4837–4838
- Geissler PL, Dellago C, Chandler D, Hutter J, Parrinello M (2000) *Chem Phys Lett* 321:225–230
- Warshel A, Levitt M (1976) *J Mol Biol* 103:227
- Field MJ, Bash PA, Karplus M (1987) *J Am Chem Soc* 109:8092–8094
- Field MJ, Bash PA, Karplus M (1990) *J Comput Chem* 11:700–733
- Monard G, Merz KMJ (1999) *Acc Chem Res* 32:904–911
- Khandogin J, Hu A, York DM (2000) *J Comput Chem* 21:1562–1571
- Van Der Vaart A, Gogonea V, Dixon SL, Merz KM Jr (2000) *J Comput Chem* 21:1494–1504
- Clark T (2000) *J Mol Struct (THEOCHEM)* 530:1–10
- Stewart JJP (1990) In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 1. VCH, New York, pp 45–81
- Zerner MC (1991) In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 2. VCH, New York, pp 313–365
- Thiel W (1997) *J Mol Struct (THEOCHEM)* 398–399:1–6
- Boyd DB (1997) *J Mol Struct (THEOCHEM)* 401:219–225
- Reynolds CH (1997) *J Mol Struct (THEOCHEM)* 401:267–277
- Dewar MJS, Zuebis EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902–3909
- Stewart JJP (1989) *J Comput Chem* 10:209–220
- Jurema MW, Shields GC (1993) *J Comput Chem* 14:89–104
- Kallies B, Mitzner R (1995) *J Mol Model* 1:68–78
- Dannenberg JJ (1997) *J Mol Struct (THEOCHEM)* 401:279–286
- Ventura ON, Coitiño EL, Lledos A, Bertran J (1989) *J Mol Struct (THEOCHEM)* 187:55–68
- Coitiño EL, Irving K, Rama J, Iglesias A, Paulino M, Ventura ON (1990) *J Mol Struct (THEOCHEM)* 210:405–426
- Khalil M, Woods RJ, Weaver DF, Smith VH Jr (1991) *J Comput Chem* 12:584–593
- Coitiño EL, Ventura ON, Sosa RM (1992) *J Mol Struct (THEOCHEM)* 254:315–328
- Cramer CJ, Truhlar DG (1994) *J Am Chem Soc* 116:3892–3900
- Rusinska-Roszak D, Lozynski M, Mack H-G, Korn M (1995) *J Mol Struct (THEOCHEM)* 342:33–41

29. Morpurgo S, Bossa M, Morpurgo GO (1998) *J Mol Struct (THEOCHEM)* 429:71–80
30. Trong-Anh N, Frison G, Solladie-Cavallo A, Metzner P (1998) *Tetrahedron* 54:12841–12852
31. Cummins PL, Titmuss SJ, Jayatilaka D, Bliznyuk AA, Rendell AP, Gready JE (2002) *Chem Phys Lett* 352:245–251
32. Spanget-Larsen J (1997) *Theor Chem Acc* 98:137–153
33. Buss V, Messinger J, Heuser N (1991) *QCPE Bull* 11:5–7
34. Csonka GI (1993) *J Comput Chem* 14:895–898
35. Cativiela C, Dillet V, Garcia JJ, Mayoral JA, Ruiz-López MF, Salvatella L (1995) *J Mol Struct (THEOCHEM)* 331:37–50
36. Csonka GI, Angyan JG (1997) *J Mol Struct (THEOCHEM)* 393:31–38
37. Salvatella L, Mokrane A, Cartier A, Ruiz-López MF (1998) *Chem Phys Lett* 296:239–244
38. Gonzalez-Lafont A, Truong TN, Truhlar DG (1991) *J Phys Chem* 95:4618–4627
39. Rossi I, Truhlar DG (1995) *Chem Phys Lett* 233:231–236
40. Geudtner J, Jug G (1993) *J Comput Chem* 14:639–646
41. Svanberg M, Pettersson JBC, Bolton K (2000) *J Phys Chem A* 104:5787–5798
42. Andersen KB, Spanget-Larsen J (1997) *Spectrochim Acta Part A* 53:2615–2625
43. Burstein KY, Isaev AN (1984) *Theor Chim Acta* 64:397–401
44. Bliznyuk AA, Voityuk AA (1988) *J Mol Struct (THEOCHEM)* 164:343–349
45. Martin B (2003) PhD thesis. Universität Erlangen-Nürnberg
46. Repasky MP, Chandrasekhar J, Jorgensen WL (2002) *J Comput Chem* 23:1601–1622
47. Bernal-Uruchurtu MI, Ruiz-López MF (2000) *Chem Phys Lett* 330:118–124
48. Bernal-Uruchurtu MI, Martins-Costa MTC, Millot C, Ruiz-López MF (2000) *J Comput Chem* 21:572–581
49. Dewar MJS, Walter T (1977) *J Am Chem Soc* 99:4899–4907
50. Halkier A, Koch H, Jorgensen P, Christiansen O, Beck Nielsen IM, Helgaker T (1997) *Theor Chem Acc* 97:150–157
51. Saint-Martin H, Hernandez-Cobos J, Bernal-Uruchurtu MI, Ortega-Blake I, Berendsen HJC (2000) *J Chem Phys* 113:10899–10912
52. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Ennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (2001) *Gaussian 98, revision A.11. Gaussian, Pittsburgh, PA*
53. (a) Rinaldi D, Hoggan PE, Cartier A (1989) *QCPE* 584a; (b) Rinaldi D, Hoggan PE, Cartier A (1989) *QCPE Bull* 128